



VALIDATION OF A TEST MEASURING YOUNG LEARNERS' GENERAL L2 ENGLISH VOCABULARY KNOWLEDGE

Liss Kerstin SYLVÉN¹ & Pia SUNDQVIST²

Abstract: This study aims to present the validation of a test designed to assess young learners' general L2 English vocabulary knowledge, the Young Learner Vocabulary Assessment Test (YLVAT). YLVAT consists of 37 items selected from the K1–2 frequency levels of the Productive and Vocabulary Levels Tests. In the study, Swedish learners ($N = 52$, age 12) took YLVAT and filled out an evaluation; scores from the national test of English (reading and listening comprehension) were also collected. Four validity measurements were used: the spread of YLVAT scores, correlation with the national test – reading, correlation with the national test – listening, and evaluation responses. YLVAT results point to a sufficient spread of scores ($M = 18.9$, $SD = 6.5$). There were significant correlations between YLVAT and (i) reading ($r = .597^{**}$), (ii) listening ($r = .541^{**}$), (iii) perceived test difficulty ($r_s = -.538^{**}$), and (iv) how fun it was to take the test ($r_s = .683^{**}$). An ANOVA showed that learners who found YLVAT “easy” or “very easy” scored significantly ($p = .000$) higher (26.1) than those who found it “difficult” (18.3) or “very difficult” (12.9). Finally, teachers found YLVAT to correlate with their own assessment of learners' vocabulary.

Keywords: young language learners, L2 English, vocabulary proficiency, formative assessment

Özet: Bu çalışma, küçük yaştaki öğrencilerin ikinci dil olarak kelime bilgilerini değerlendirmek için tasarlanmış bir testin, Küçük Yaştaki Öğrencilerin Kelime Değerlendirme Testinin (YLVAT), geçerliliğini sunmayı amaçlanmaktadır. YLVAT, Kelime Seviyeleri Testlerinin K1-2 sıklık seviyelerinden seçilmiş olan 37 maddeden oluşmaktadır. Bu çalışmada, İsveçli öğrenciler (sayı=52, yaş 12) YLVAT'yi aldılar ve bir değerlendirme formu doldurdular. Aynı zamanda, ulusal İngilizce testinin (okuma ve dinlemediğini anlama) sonuçları da toplandı. Dört adet geçerlilik ölçümü kullanıldı; YLVAT puanlarının yayılımı, ulusal test ulusal test –okuma– ile korelasyon okuma- ile korelasyon, ulusal test –dinleme– ile korelasyon, değerlendirmedeki cevaplar. YLVAT sonuçları, puanları yeterli derecede dağılımına dikkat çekmektedir ($M = 18.9$, $SD = 6.5$). YLVAT ve (i) okuma ($r = .597^{**}$), (ii) dinleme ($r = .541^{**}$), (iii) algılanan test zorluğu ($r_s = -.538^{**}$), ve (iv) teste girmenin ne kadar eğlenceli olduğu ($r_s = .683^{**}$) arasında önemli korelasyonlar bulunmuştur. ANOVA göstermiştir ki YLVAT'ı ‘kolay’ veya ‘çok kolay’ bulanlar, onu ‘zor’ (18.3) veya ‘çok zor’ (12.9) bulanlardan anlamlı derecede ($p = .000$) daha fazla puan (26.1) almışlardır. Sonuç olarak, öğretmenler YLVAT'ı öğrencilerin kelime bilgileri ile alakalı kendi değerlendirmeleri ile ilişkili bulmuşlardır.

Anahtar sözcükler: küçük yaştaki dil öğrencileri, ikinci dil olarak İngilizce, kelime bilgisi yeterliliği, biçimlendirici değerlendirme

1. Introduction

Traditionally, language learning of younger and older adults has been favored over that of children in research (for an overview, see Hasselgreen & Drew, 2012; M. Nikolov, 2009). However, over the last decade, something has happened within the broad field of second language acquisition (SLA). The sheer number of volumes published with a focus on

¹ PhD, English Linguistics, Department of Education and Special Education, University of Gothenburg, Sweden, Lisskerstin.sylvén@ped.gu.se

² PhD, English Linguistics, Department of English, Faculty of Arts and Social Sciences, Karlstad University, Sweden, Pia.sundqvist@kau.se

foreign/second language (L2) learning among young learners has grown (see, e.g., García Mayo & García Lecumberri, 2003; Hasselgreen, Drew, & Sørheim, 2012; Marianne Nikolov, 2009), and also large transnational research projects targeting young language learners have been conducted (see, e.g., Enever, 2011). This broadened scope which includes young learners in SLA research is indeed welcome.

As known, formative assessment is generally preferred for young language learners (YLLs) (McKay, 2006). When working with YLLs, it is particularly important to consider cognitive, social, emotional, and physical growth in relation to vulnerability. Therefore, it is generally suggested that summative assessment should not be used for L2 language learning in primary school; instead, it should be introduced at a later stage (McKay, 2006), which often tends to be the case. However, it is worth mentioning that the introduction of formal assessment of foreign language abilities varies among countries. In Sweden, for instance, grades and national tests in English are used for the first time in 6th grade, at the age of 12 (The Swedish National Agency for Education, 2016). In contrast, for children in the Netherlands – a seemingly similar country in many respects – the abilities in various school subjects, including English, are already assessed annually from the beginning of primary school (The Government of the Netherlands, 2016). Thus, from an international perspective, we see different practices as regards the formal grading/assessment of foreign languages. In this paper, we introduce the Young Learner Vocabulary Assessment Test (YLVAT), a vocabulary test designed for young English language learners. As is clear in the overview below of existing tests for YLLs, vocabulary tests are few and far between. YLVAT is to be of formative or diagnostic use, facilitating a teacher's initiation of individualized intentional L2 English vocabulary learning among YLLs. The main aim of this paper is to validate YLVAT in order to establish whether it may be suitable for measuring general vocabulary knowledge among YLLs. A secondary aim is to investigate young test-takers' evaluation of YLVAT.

2. Theoretical background

2.1. Knowing a word and assessing vocabulary knowledge

When vocabulary knowledge is focused, the first concern is to define what “knowing a word” actually means. The second concern is how vocabulary knowledge can be assessed. To start with the first question, there is no clear-cut answer. Several attempts have been made at listing various characteristics that need to be taken into account in order to be able to claim knowledge of a word (see Carter, 1987; Meara, 1996). The list by Nation (1990) includes knowing the meaning(s), the spelling, the pronunciation, the semantic properties, and the collocational patterns of the word. These five specific aspects are also the characteristics that stand out as being generally agreed upon as central (see, e.g., Henriksen, 1999; Paribakht & Wesche, 1997). For the purpose of this paper, where a written test is used (for a validation of an aural vocabulary test, see McLean, Kramer, and Beglar, 2015), the meaning and the collocational patterns of a word are the core characteristics investigated.

The second question is how vocabulary knowledge can be assessed. The multiple-choice format is often used for vocabulary testing. It typically consists of a number of alternative options (normally 3–5) from which the correct answer is to be chosen. There are many benefits of the multiple-choice format. For instance, it is easily administered and quickly graded, and there is only one possible answer. However, one drawback is that it requires a great deal of work in order to construct suitable distractors. The *Vocabulary Levels Test*, VLT, (Laufer & Nation, 1999;

Nation, 2001) is another test format, originally constructed for diagnostic purposes among adults, but also used in many other contexts (Beglar, 2010). In the VLT, six items are listed, out of which three are to be paired up with synonyms or explanations offered. Again, this is an easy test to administer and grade, but unless online tests are used, the construction work is laborious. Yet another type of test is the *yes/no-test* (Meara, 1992), where test-takers indicate whether a certain word is known (yes) or unknown (no). A number of nonsense words are included to ascertain that test-takers are trustworthy when answering. This, too, is an easily administered test, but a great disadvantage is that sophisticated analyses are needed to correctly control for the nonsense words and the answers given for them (Mochida & Harrington, 2006). All tests mentioned so far tap into receptive vocabulary knowledge; that is, they focus on the meaning of the target item. In contrast, the *Productive Levels Test*, PLT, taps into productive vocabulary knowledge. In the PLT (Laufer & Nation, 1999; see also, Lextutor), test-takers are supposed to produce a target word, for which the first few letters are given. A one-sentence context is provided: “He was riding a bi.....” [*bicycle*]. The PLT stems from the so-called *C-test*, which has been shown to be a highly reliable and valid test based on the principle of reduction of redundancy (see, e.g., Klein-Braley, 1996; Huhta, 1996). The PLT/C-test format has been criticized for being limited, tapping only into lower-level processing in language performance, but Klein-Braley (1996) argues convincingly for the opposite, claiming that test performance “can validly be interpreted in terms of general or overall language proficiency” (p. 91). She also shows that the test format can be used with young learners. Further, Huhta’s (1996, p. 217) results, from a study involving EFL learners, revealed that test-takers found the gap-filling format “interesting” in comparison with other test formats.

A test that can be used for both receptive and productive vocabulary knowledge, depending on the design, is a so-called *cloze-test*. The cloze-test is characterized by a certain number of words being left out of a text. It can be designed so that every n^{th} word is deleted, or selected words can be left out. Thus, it is a very flexible type of test, capable of being tailored for many purposes. Further, the cloze-test can be designed to tap into either receptive and/or productive word knowledge, and as for productive knowledge, the meaning of the word, spelling, and collocational patterns can all be tested.

2.3. Vocabulary tests for YLLs

As research has shown, young EFL learners in different countries encounter English not only in school, but also to a large degree outside of school (Forsman, 2004; Henry, 2013; Olsson, 2011; Sundqvist, 2011; Sundqvist & Sylvén, 2012, 2014; Sylvén & Sundqvist, 2012). In line with Sundqvist (2011), we refer to such out-of-school contacts as *extramural English* (EE), which comes in many forms and shapes, very much depending on national context. In the Nordic countries, the Netherlands, and Belgium, for instance, societal EE abounds with subtitled, rather than dubbed English-speaking TV-shows and films; English words and phrases are being used in ads as well as in every-day language, etcetera. This can be compared with Spain and Italy, where English is rather sparse in society and dubbing is commonly used for English productions. Another characteristic of EE is that it is also very private; English use in individuals’ social networks as a communicative tool is predominant, even among EFL learners as young as ten-year-olds (Sundqvist & Sylvén, 2014). In a decade-old nationwide Swedish investigation, more than half of the 5th graders claimed that they learn as much or more English outside of school than in school (The Swedish National Agency for Education, 2004). The fact that the amount and type of EE varies immensely between individuals, also at a very young age, means that the type

of vocabulary acquired through EE also differs at an individual YLL level. One way of addressing this is suggested by Uzun (2009), who makes an attempt to introduce digital games for foreign language learning in the classroom. The items chosen for YLVAT, therefore, represent general vocabulary, rather than subject-specific vocabulary that may have been taught in school. What YLVAT intends to measure is the overall level of vocabulary knowledge among YLLs, and not specifically to check what has been taught and learned in school by the learners. An existing vocabulary test that may also be used for YLLs is the *Peabody Picture Vocabulary Test* (Dunn & Dunn, 2007). However, it is deemed too simplistic for YLLs in countries in which the presence of English is substantial (cf., Berns, de Bot, & Hasebrink, 2007). Moreover, Cambridge English Language Assessment offers three types of English tests called *Starters*, *Movers*, and *Flyers* (Cambridge English), but in these several language abilities other than vocabulary are tested, such as listening, reading and writing, and speaking.

As pointed out by Scott Langeland (2012), it is important for teachers to learn about their own students' vocabulary knowledge, since research shows that middle school children whose L2 vocabulary is small are at risk in terms of L2 development (Roessingh & Elgie, 2009). Thus, the possibility of using a vocabulary test could constitute a helpful tool for teachers in their daily work. However, L2 English vocabulary tests designed for YLLs are few and far between, even though, as indicated by the recent work at Cambridge mentioned above, there is a rising interest in constructing valid and reliable tests for this category of learners.

In an evaluation about views on language testing and assessment among teachers and students in compulsory schools in ten European countries, Erickson and Gustafsson (2005) show that these young students indeed often see tests as a learning opportunity, rather than something they wish did not exist. Also, even though arguably many young test-takers only want to be tested on what they have learned in school, others feel that “[...] a good test should not be too easy” (Erickson & Gustafsson, 2005: 13, original spelling). In other words, tests can and should include items across the entire continuum, from easy, frequent words to difficult, more infrequent ones.

A question remains, however, with regard to what might constitute an “average” size of L2 English vocabulary among YLLs today. It seems difficult to estimate the average size of a YLL's L2 English vocabulary, and it is highly dependent on the national context. Nation (2001) says that in countries where English is taught as a foreign language, after taking English in school for about 40 weeks per year for five years, learners would be expected to be familiar with high frequency words in English, and at best, know approximately 2,000 word families. Laufer (1998) found that 10th-grade students in Israel averaged about 1,900 word families after six years of instruction. Hasselgreen (1996) estimated Norwegian 8th-grade students to have a passive vocabulary knowledge of about 1,600 word families. With these studies as a backdrop for creating a test of general L2 English vocabulary for learners at the end of primary school, it seems suitable to include target words from the 1,000 and 2,000 frequency levels.

At this point, we would like to comment briefly on the fact that we, in the results section below, distinguish between genders. This distinction is made because earlier findings on EE and L2 vocabulary indicate significant gender differences, namely that boys have a larger L2 English vocabulary than girls (Olsson, 2011; Sundqvist, 2011; Sundqvist & Sylvén, 2012; Sylvén & Sundqvist, 2012). Other studies, not dealing specifically with EE and/or YLLs, also indicate that

boys are at an advantage as regards vocabulary proficiency (Herriman, 1997; Reuterberg, 1999). Our hypothesis, therefore, is that in YLVAT, boys will also score higher.

2.4. Young Language Vocabulary Assessment Test

As mentioned above, the aim of YLVAT is to tap into students' general vocabulary knowledge, rather than vocabulary relating to any specific subject domain. The items included were selected from existing VLTs and PLTs (Lextutor) and, apart from representing general vocabulary, a number of principles were used. First of all, items were selected from the 1,000 and 2,000 levels. This delimitation is due to the reasonable assumption that the vocabulary knowledge of YLLs does not extend beyond these two levels (cf. above), and to maintain an acceptable level of difficulty for test-takers. Furthermore, the forms of words chosen varied in number: the majority were nouns, some were verbs, and a few were adjectives. Finally, some cognates (for an overview, see Rogers, Webb, & Nakata, 2015) were included among the items tested; in tests for YLLs, it is of the essence not to crush their self-confidence (McKay, 2006), and including cognates was mainly done as an affective measure, as we anticipated that most learners would be able to recall or reproduce such items in English.

When designing a test for YLLs, the risk of test fatigue needs to be seriously considered. For this reason, together with other reasons such as test validity, it was decided to include three different test formats in YLVAT. The first part, Part A, consists of thirteen statements to which the test-taker has to indicate whether they are a) true (T), b) not true (N), or c) not known (X). An example from this part is:

- 1) All the world is under water. ☐ T
- ☐ N
- ☐ X

As is clear, Part A tests word recognition. All items are taken from the 1,000 level.

The second part of YLVAT, Part B, consists of twelve items from the VLT, as described above. An example from this part is:

- Question 1) 1. Apply
2. Elect -- choose by voting
3. Jump -- become like water
4. manufacture -- Make
5. Melt
6. Threaten

The items in Part B are taken from the 2,000 level, with six nouns and six verbs.

Finally, the third part, Part C, consists of 12 sentences chosen from the PLT. In each sentence, one word is left out, but the initial two or three letters are given, as in the following example:

- 1) Plants receive water from the soil through their ro_____.

Thus, Part C taps into students' productive vocabulary knowledge, as opposed to Parts A and B, where receptive knowledge is in focus. The target items are from the 2,000 level and consist of five nouns, four verbs, and three adjectives.

In sum, YLVAT represents widely recognized test types. It taps into receptive knowledge of the meanings of words in Parts A and B, as well as productive knowledge, certain syntactic and collocational patterns, and knowledge of spelling in Part C. For a full version of YLVAT, see Appendix 1.

3. Aim and research questions

The overarching aim of this paper is to evaluate and validate YLVAT by investigating whether it is a reliable tool for measuring YLLs' general L2 English vocabulary knowledge. In order to achieve the aim, five research questions guided the present study:

1. Does YLVAT yield a normal distribution of scores?
2. Are there gender differences in YLVAT scores?
3. Does YLVAT correlate with national test scores on listening and reading comprehension?
4. What do test-takers think about YLVAT as regards (a) level of difficulty and (b) degree of funniness?
5. What do teachers of young L2 English learners think about YLVAT?

4. Method

4.1. Participants

The participants are from three classes at two municipal schools in a medium-sized town in Sweden, and they took YLVAT early in the fall semester of 6th grade (age 12). For the study presented here, it is important to report the total number of hours of formal English instruction the participants had had prior to taking YLVAT, and it is estimated to be approximately 300 (c. 12 hours in 1st and 2nd grade; 23 in 3rd; 120 in 4th; 131 in 5th).

In total, 52 students participated, 31 girls and 21 boys. Thus, the sample is fairly small, which calls for caution when interpreting the results. We will return to this in our discussion. In the sample, almost all (48 students, or 92%) spoke Swedish as a first language (L1). The remaining four students were bilinguals (Swedish-German/Persian/Russian/Spanish). In terms of language background, our sample differs somewhat from Swedish compulsory school in general, where approximately 20% of the students have another L1 than Swedish (The Swedish National Agency for Education, 2016). Nevertheless, we would like to describe the participating learners as typical, 12-year-old boys and girls from an ordinary Swedish town, who thus constitute a suitable sample for validating and evaluating a test such as YLVAT.

4.2. Material and test administration

Copies of YLVAT were distributed to the teachers along with careful oral and written instructions about how to administer the test in the classroom. The maximum time was set to 40 minutes. The teachers were encouraged to note how long it took for each student to complete the test, and for most of them, about 15 minutes was enough. As can be seen in Appendix 1, YLVAT

includes dual-language instructions, Swedish (the L1) plus English (the target language). By using both languages in the written test instructions, the need for additional oral instructions was deemed small, but if asked by a participant, the teachers were instructed that they were allowed to pronounce particular words aloud.

On the actual test occasion, the participants were informed by their teacher that they were to take a new vocabulary test, and that two researchers (one of whom they had previously met) *wanted to test if it was suitable for use in 6th grade*. For this reason, the participants were told that there was also a sheet of evaluation attached to the test, where they were encouraged to respond as sincerely as possible to four questions: (1) Which part of the test was the easiest? (2) Which part of the test was the most difficult? (3) How difficult was the test overall? (very difficult/difficult/easy/very easy) (4) How fun was it to take the test? (very fun/fun/boring/very boring). Evaluation responses were also used as part of the test validation. The participants were also informed that the researchers were to correct the tests, and that the tests would be returned to the students later, along with test scores to the teacher; test copies were to be kept by the researchers. The teacher then read the introduction to the test aloud and demonstrated the Part A example on the whiteboard, leaving room for the participants to ask questions before they commenced taking the test.

When YLVAT had been corrected and returned, as another part of the validation process, the teachers ($n = 2$) were interviewed about the instructions and the test, and asked to what extent they thought the test scores matched their own opinions on their students' level of vocabulary knowledge (cf. Read & von Randow, 2013, about collecting feedback from both students and teachers for the purpose of language test validation). In addition, copies of learner scores on the national test of English for the participating classes were collected from the teachers. Results from the reading and listening comprehension parts were used in correlation analyses with test scores to validate YLVAT. A methodological reason for making such correlations is based on findings from previous research on so-called lexical coverage (the proportion of known words in a specific piece of discourse, Adolphs & Schmitt, 2003), which shows links between vocabulary size and listening as well as reading comprehension (see, e.g., Laufer, 1995; van Zeeland & Schmitt, 2013; Webb & Rodgers, 2009).

4.3. Correcting

All tests were corrected by the researchers and double-checked. For purposes of statistical analyses, for Part A, participant responses were first inputted into statistical software and then an additional variable was created in which we inputted for each case (participant) whether s/he actually knew the answer or not. For Part B, the statistical procedure was repeated in a similar fashion. However, for Part C, the productive part where the participants should write a target word, in addition to inputting whether the word was known or not, we also created a variable coded for (i) no word provided, (ii) incorrect word provided, (iii) correct word and spelling, and (iv) correct word but incorrect spelling.

4.4. Analytic procedure

The research questions ask whether YLVAT is a valid tool for measuring vocabulary knowledge among young learners of English. Validity has been defined as “an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of

assessment” (Messick, 1989, p.13, italics in the original). For the purpose of validating YLVAT, four measurements were used: (i) the spread of scores, (ii) correlation with scores on the national test of reading comprehension, (iii) correlation with scores on the national test of listening comprehension, and (iv) participant responses to the evaluation of YLVAT. An additional research question has to do with gender-related test differences. We use a quantitative method of analysis to answer all these questions, whereas teacher interview data were analyzed qualitatively.

All statistical tests were run in IBM SPSS Statistics 22. We regard $p < .05$ as significant and report exact p -values to facilitate interpretation. One-way analysis of variance (ANOVA) together with classical eta squared (η^2) were used to calculate significance and effect sizes. Cohen’s conventions for interpreting effect sizes were used (see Aron, Aron, & Coups, 2005). In line with Cohen’s convention for r^2 , $\eta^2 = .01$ is a small effect size, $\eta^2 = .06$ is medium, and $\eta^2 = .14$ is large (Dörnyei, 2007, p. 221). For ANOVA, Gabriel’s post-hoc test, suitable for unevenly-sized groups, was used to provide additional indications of which groups differed from which within the general differences between groups. The independent samples t -test was used for group comparisons (e.g., boys and girls). For correlation analyses, Pearson’s product-moment correlation coefficient (r) and Spearman (r_s) were used.

5. Results

In this section, the results obtained on the YLVAT are presented. First, the overall findings are presented including the results for each part of the test, and then an item-level analysis is given.

5.1. Overall results

In Figure 1, the overall results are illustrated in the form of a histogram.

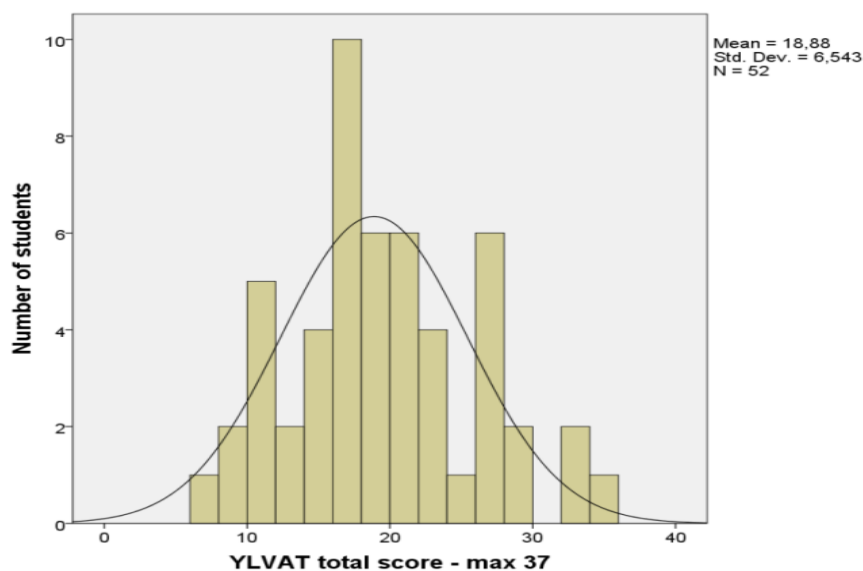


Figure 1. Total score.

As shown in Figure 1, the distribution of total scores is normal, demonstrating that the test discriminates well between test-takers. The total maximum score is 37, which was not obtained by anybody, and the mean score was close to 19. The highest score obtained was 34, and the

lowest score was 7, indicating that all test takers were able to get 7 points or more. In Table 1, the results are shown, divided up between boys and girls.

Table 1. Total scores, boys vs. girls

	Boys (<i>n</i> = 21)	Girls (<i>n</i> = 31)	<i>p</i> -value
Mean	21.3	17.2	.025
SD	6.6	6.1	
Min	10.0	7.0	
Max	34.0	32.0	

In Table 1, we see that the boys' score was significantly higher than the girls' ($p = .025$). Further, the boys have a higher minimum as well as maximum score, and their SD is slightly larger than that of the girls. Table 2 illustrates the scores on Part A per gender.

Table 2. Scores on part A, boys vs. girls

	Boys (<i>n</i> = 21)	Girls (<i>n</i> = 31)	<i>p</i> -value
Mean	10.0	9.4	.285
SD	2.1	2.1	
Min	7.0	5.0	
Max	13.0	13.0	

Table 2 reveals that boys and girls perform in a similar manner on Part A, which, as explained above, is the easiest part of YLVAT. No statistically significant difference was found. The mean is high for both boys and girls, and several students scored the maximum of 13 points on this part.

Part B taps into receptive vocabulary taken from the VLT, where three out of six distractors are to be paired up with a synonym or paraphrase. This part contains twelve items at the 2,000 level, and the results are illustrated in Table 3.

Table 3. Scores on Part B, boys vs. girls

	Boys (<i>n</i> = 21)	Girls (<i>n</i> = 31)	<i>p</i> -value
Mean	6.3	4.2	.008
SD	2.7	2.8	
Min	1.0	0.0	
Max	11.0	10.0	

As shown in Table 3, the boys outperform the girls significantly ($p = .008$) on Part B, and possible reasons for this are elaborated on below.

Finally, Part C is the part of YLVAT that taps into productive vocabulary knowledge (12 sentences with one word/sentence to be completed), and the results are presented in Table 4.

Table 4. Results on Part C, boys vs. girls

	Boys ($n = 21$)	Girls ($n = 31$)	p -value
Mean	5.0	3.7	.060
SD	2.9	2.1	
Min	0	0	
Max	11.0	9.0	

As can be seen in Table 4, the boys score higher than the girls, but not significantly so ($p = .060$). As indicated by the mean scores, Part C was difficult for both boys and girls, and some students scored zero. In the following section, we give a detailed analysis of each item in the three parts.

5.2. Item-level

In this section, the solution pattern for each part of the test is analyzed in total as well as from a gender perspective. First, in Table 5, the 13 items in Part A are listed along with the percentage of correct and incorrect answers.

Table 5. Scores, Part A, total and per gender

	Total ($n = 52$)		Boys ($n = 21$)		Girls ($n = 31$)	
Item	Wrong	Right	Wrong	Right	Wrong	Right
1	1.9	98.1	0.0	100.0	3.2	96.8
2	38.5	61.5	42.9	57.1	42.9	57.1
3	7.7	92.3	9.5	90.5	9.5	90.5
4	3.8	96.2	0.0	100.0	6.5	93.5
5	32.7	67.3	23.8	76.2	38.7	61.3
6	3.8	96.2	0.0	100.0	6.5	93.5
7	21.2	78.8	9.5	90.5	29.0	71.0
8	61.5	38.5	47.6	52.4	71.0	29.0
9	36.5	63.5	47.6	52.4	29.0	71.0
10	19.2	80.8	19.0	81.0	19.4	80.6

11	48.1	51.9	38.1	61.9	54.8	45.2
12	15.4	84.6	19.0	81.0	12.9	87.1
13	48.1	51.9	42.9	57.1	51.6	48.4

In Table 5, it is shown that statement 8 (*Remain here means stay*) proved to be the most difficult. Other difficult items include 11 (*Each society has the same rules*) and 13 (*It is a short way from one side to the other side of a wide river*). Item 13 is a very long sentence, where the target word “wide” comes second to last. There were three items on which boys were 100% correct, namely item 1 (*Two of these are little*), 4 (*All the world is under water*), and 6 (*Sometimes people die when they fall off a building*). These were also the easiest items for the girls. The results of an item-level analysis for Part B are given in Table 6.

Table 6. Scores, Part B, total and per gender

Item	Total (N = 52)			Boys (n = 21)			Girls (n = 31)		
	No answer	Right	Wrong	No answer	Right	Wrong	No answer	Right	Wrong
Q1a <i>elect</i>	26.9	38.5	34.6	38.1	38.1	23.8	19.4	38.7	41.9
Q1b <i>melt</i>	17.3	65.4	17.3	19.0	76.2	4.8	16.1	58.1	25.8
Q1c <i>manufacture</i>	50.0	13.5	36.5	66.7	9.5	23.8	38.7	16.1	45.2
Q2a <i>hide</i>	28.8	58.8	15.4	19.0	76.2	4.8	35.5	41.9	22.6
Q2b <i>spoil</i>	50.0	15.4	34.6	61.9	19.0	19.0	41.9	12.9	45.2
Q2c <i>invite</i>	23.1	63.5	13.5	14.3	81.0	4.8	29.0	51.6	19.4
Q3a <i>pride</i>	30.8	40.4	28.8	33.3	47.6	19.0	29.0	35.5	35.5
Q3b <i>debt</i>	48.1	23.1	28.8	52.4	23.8	23.8	45.2	22.6	32.3
Q3c <i>roar</i>	34.6	42.3	23.1	33.3	61.9	4.8	35.5	29.0	35.5
Q4a <i>salary</i>	42.3	30.8	26.9	42.9	38.1	19.0	41.9	25.8	32.3
Q4b <i>temperature</i>	21.2	61.5	17.3	9.5	81.0	9.5	29.0	48.4	22.5
Q4c <i>flesh</i>	32.7	46.2	21.2	23.8	66.7	9.5	38.7	32.3	29.0

As shown in Table 3, boys scored significantly higher than girls on Part B in general. Table 6 reveals that the words *hide*, *invite*, *melt*, and *temperature* were the easiest for all test-takers, but in particular for the boys. On the other hand, the words *debt*, *manufacture*, and *spoil* proved to be the most difficult, and approximately 50% of the test-takers did not give an answer at all to these.

The words *invite* (Swe. ‘invitera’) and *temperature* (Swe. ‘temperatur’) are English-Swedish cognates, and approximately 60% of the test-takers were correct on these items. *Melt* is another word where the majority answered correctly, which was also the case for *hide* (c. 60% correct answers). Some gender differences can be found in Part B: *flesh*, *hide*, *meat*, *melt*, and *roar* have a higher solution rate for the boys than the girls. Table 7 gives the scores for Part C.

Table 7. Scores, Part C, total and per gender

Item	Total (N = 52)		Boys (n = 21)		Girls (n = 31)	
	Wrong	Right	Wrong	Right	Wrong	Right
1: <i>roots</i>	61.5	38.5	47.6	52.4	71.0	29.0
2: <i>nurse</i>	42.3	57.7	42.9	57.1	41.9	58.1
3: <i>tip</i>	53.8	46.2	33.3	66.7	67.7	32.3
4: <i>motor</i>	21.2	78.8	14.3	85.7	25.8	74.2
5: <i>copy</i>	71.2	28.8	66.7	33.3	74.2	25.8
6: <i>climb</i>	34.6	65.4	23.8	76.2	41.9	58.1
7: <i>connects</i>	96.2	3.8	100.0	0.0	93.5	6.5
8: <i>surrounded</i>	86.5	13.5	76.2	23.8	93.5	6.5
9: <i>usual</i>	88.5	11.5	76.2	23.8	96.8	3.2
10: <i>wandered</i>	4.6	15.4	81.0	19.0	87.1	12.9
11: <i>hungry</i>	34.6	65.4	28.6	71.4	38.7	61.3
12: <i>examined</i>	98.1	1.9	95.2	4.8	100.0	0.0

As illustrated in Table 7, there are some easy and some difficult words also in Part C. The easy ones are *climb*, *hungry*, and *motor*. *Hungry* and *motor* are Swedish cognates, and *climb* is a high-frequency motion verb commonly used in primary school classrooms (Lextutor). A word of moderate difficulty is *tip*, for which 68% of the girls answered correctly, as compared with 33% for the boys. The most difficult words were *connects*, *examined*, *surrounded*, and *usual*; all appear at the very end of the test. The participle form *surrounded* is undeniably a difficult word, and the most difficult word of all was *examined*, which only one learner (a boy) knew.

Before moving on to the results of the correlations of YLVAT scores and national test scores, we would like to conclude this section by mentioning the fact that no gender difference as regards the amount of effort put into doing the test was found: the boys and the girls seem to have made an equal effort based on the ratio of left-out answers.

5.3. Correlations with national test scores

For the purpose of validating YLVAT, the test scores were correlated (Pearson, two-tailed) with the learners' scores on two parts (reading and listening comprehension) of the mandatory national test of English used in Sweden. The sample had a mean score of 15.7 (SD = 3.3) on the reading test (max = 19) and a mean of 26.2 (SD = 6.0) on the listening test (max = 31). The results revealed statistically significant correlations between YLVAT and reading comprehension ($r = .597$, $p = .000$) as well as between YLVAT and listening comprehension ($r = .541$, $p = .000$).

5.4. Student evaluation

Based on the responses to question 1 in the evaluation, the vast majority of the test-takers ($n = 47$) considered Part A to be the easiest part (3 missing). With regard to which part was the most difficult, 36 participants answered Part C, whereas 13 said Part B (3 missing). Question 3 asked about the overall difficulty of YLVAT; the results are presented in Table 8, which shows that most test-takers found the test difficult or very difficult (86%); no one found it very easy. The mean YLVAT scores for three groups based on the variable "perceived test difficulty" (easy/difficult/very difficult) are shown in Table 9.

Table 8. Responses to question 3 (perceived test difficulty)

	Very easy	Easy	Difficult	Very difficult
How difficult was the test overall? (N = 48)	0	7	33	8
Responses (%)	0	14	69	17

Table 9. Perceived test difficulty and YLVAT scores

Perceived test difficulty	N	YL VAT score (M)	SD
Easy	8	26.1	6.6
Difficult	34	18.3	5.4
Very difficult	8	12.9	4.7
Total	50	18.7	6.6

To further investigate the relation between perceived test difficulty and YLVAT scores, correlation analysis was carried out. Results revealed a significant negative correlation ($r_s = -.538$, $p = .000$), indicating that the more difficult a learner thought it was to take YLVAT, the lower the score. It also means that approximately 29% of the variance in test scores was explained by perceived test difficulty. In addition, an ANOVA revealed that the differences between the three groups were significant ($p = .000$), and Gabriel's post hoc showed that the YLLs who thought YLVAT was easy had a mean score that differed significantly from each of the other two groups (compared with "Difficult", $p = .001$; with "Very difficult", $p = .000$). The two groups, "Difficult" and "Very difficult", also differed significantly from one another in terms of YLVAT scores ($p = .030$), leading to the conclusion that all groups differed significantly from one another in terms of YLVAT scores. The effect size was large ($\eta^2 = .336$).

The fourth question ("How fun was it to take the test?") included four response options and the answers were as follows: very boring ($n = 5$), boring ($n = 28$), fun ($n = 17$), and very fun ($n = 1$) (1 missing). The two "fun-options" were collapsed into one category in the subsequent analyses, since only one had responded "very fun". Correlation analysis was used to examine whether there was a connection between how fun it was to take the test and YLVAT scores, and the results showed a significant positive correlation ($r_s = .683$, $p = .000$); that is, about 47% of the variance was explained by "the fun factor". The findings indicate that the more fun YLLs said that it was to take YLVAT, the higher they scored. Furthermore, ANOVA showed that the differences between the three groups were significant ($p = .000$), and Gabriel's post hoc revealed that the group of YLLs who thought it was fun to take the test differed significantly from "Very boring" ($p = .000$) and "Boring" (whereas there was no difference between the latter two groups). Again, the effect size was large ($\eta^2 = .393$). The mean YLVAT scores for the three fun factor groups are shown in Table 10.

Table 10. Responses to question 4 ("The fun factor")

The fun factor	N	YL VAT score (M)	SD
Very boring	5	11.6	2.7
Boring	28	16.8	5.6
Fun/Very fun	18	23.9	5.0
Total	51	18.8	6.6

Finally, Spearman correlation was then used to compare the two variables, perceived test difficulty and the fun factor: $r_s = -.556$ ($p = .000$). In sum, YLLs who responded that they thought it was fun to take YLVAT also tended to think it was easy, and vice versa.

5.5. Teacher interviews

As mentioned above, two teachers were interviewed about YLVAT, and both were positive. They thought the instructions were helpful for the actual test administration. They viewed YLVAT as an opportunity to receive additional information about individual learners' L2 English vocabulary, and appreciated the possibility of using a test for which their students had not studied in advance. When asked about their opinion on the extent to which YLVAT scores matched their own assessment of learners' L2 vocabulary, their responses were unanimous: they were happy to see how well the scores reflected their own qualitative assessment, indicating high face validity. The fact that only two teachers were involved is a limitation of this study, which we address below.

6. Discussion

In this article we have accounted for (i) the design of a vocabulary test intended for young language learners, the Young Language learner Vocabulary Assessment Test (YLVAT), and (ii) the results of this test when administered among 6th-graders. YLVAT is based on existing tests, originally designed to target adults, namely the VLT and the PLT (Laufer & Nation, 1999; Nation, 2001). By using only test items belonging to the 1,000 and 2,000 word frequency levels, and combining three different test formats, YLVAT is deemed suitable for the targeted age group of YLLs. In the following, the findings are discussed in the order of the five research questions.

The spread of scores on the test as reported above was very good, indicating that YLVAT discriminates well between learners. As expected, Part A, which above all tests word recognition, was the easiest part of the test, whereas Part C, which requires productive vocabulary knowledge, was the most difficult. Throughout the test, cognates in Swedish, such as *hungry* and *motor*, proved to be the easiest, which is also in line with what was expected. Such items were included in order to ensure the feasibility for all test takers to feel mastery of at least some target words (cf., McKay, 2006), and this intention, then, had the desirable effect. Indeed, no one scored below 7. On the other hand, in order to avoid a ceiling effect and to enable an effective discrimination among test takers, some very difficult items were also included. These are above all found in the productive part of the test, and include words such as *examined* and *surrounded*. Previous research on test takers' views on testing and assessment has found that the inclusion of difficult items or parts in a test is not necessarily apprehended as negative by test takers (cf. Erickson & Gustafsson, 2005), but that they rather appreciate a test being somewhat challenging.

Among the statements in Part A, number 8 (*Remain here means stay*) proved to be the most difficult. The reason for this is believed to be that the construction is complex, and possibly too advanced for YLLs to decipher. Other items that were difficult were items 11 (*Each society has the same rules*) and 13 (*It is a short way from one side to the other side of a wide river*). These constructions are also somewhat complex, and in item 11, the word *society* may be too abstract for learners at the age of 12. Item 13 is a very long sentence, where the target word *wide* comes second to last. The length of the sentence puts great demands on working memory, which might explain the low ratio of successful answers.

In Part B, the test becomes slightly more difficult as it contains words from the 2,000 level. The cognates *invite* and *temperature* were, as expected, easier than some other items in this part. Approximately 60% of the test takers were correct on these items. *Melt* is another item where the majority got the correct answer, which may be explained by the fact that it is connected to (the melting) snow in Sweden, most probably a popular topic in the EFL classroom at this learning stage. *Hide* is another item with about 60% correct answers, possibly thanks to the age-related collocation *hide and seek*.

Likewise, there are some easy and some difficult words also in Part C. The easy ones are *climb*, *hungry*, and *motor*. *Hungry* and *motor* are cognates, and *climb* is a motion verb appropriate for the age group in focus. With regard to *tip*, there is a gender-related difference with as many as 68% of correct answers from the girls, but only 33% from the boys, the explanation for this remains unclear.

In the sentence where *connects* appears (which 6.5% of the girls but none of the boys knew), the word *suburbs* also occurs:

The railway con_____ London with its suburbs.

If *suburbs* is unknown, the target word is difficult to figure out. Moreover, the item *usual* is used in an unexpected construction:

This work is not up to your us_____ standard

This may be too advanced for most of the test-takers. The participle form *surrounded* is undeniably a difficult word, and the most difficult word was *examined*, which few knew. *Examined* is also the very last test item and thus test fatigue most probably contributes to the low score.

As shown throughout the results section, there are some noteworthy gender differences, and, thus, the answer to our second RQ is “yes”. Part B is easier for the boys, and one reason may be the test format. It could also be that the items are more suited to boys. For instance, the words *hide*, *flesh*, *meat*, *melt*, and *roar* are much easier for the boys than the girls. We speculate that one possible explanation is these words occur in the digital games many boys in this age group engage in (Sundqvist & Sylvé, 2012).

There were statistically significant positive correlations between YLVAT, the reading comprehension part of the national test of English and the listening comprehension part of the same test. This means that there is a tendency for YLLs with good comprehension skills to also possess a substantial number of words. Bearing in mind, the main aim of this study – to validate YLVAT as a test of general L2 English vocabulary knowledge – the findings from the correlation analyses are important. Knowledge of many words is a prerequisite for good comprehension (see, e.g., Webb, 2008) and, therefore, we interpret the results of the correlation analyses as evidence of YLVAT being a valid test.

The YLLs in our study had different views about YLVAT. Clearly, many of them found the test to be difficult, which is probably the reason why many also said it was boring or very boring. Nevertheless, as many as 18 (35%) actually thought it was fun or very fun to take YLVAT. Most likely, these YLLs appreciated the fact that they had the opportunity to do something different as

compared with everyday lesson work. It is also possible that some L2 learners prefer taking tests that do not demand any prior studying. The results revealed that the fun factor explained 47% of the variance in scores, whereas perceived test difficulty explained a lower percentage (29%). In light of these findings and the results of the correlation analyses, although speculative, the fun factor can be viewed as a good predictor of the total YLVAT score, bearing in mind that perceived level of difficulty also seems to function well as a predictor.

There were only two teachers involved in this study, so the qualitative analysis of their views on YLVAT undoubtedly becomes very limited. Even so, for the purpose of the study, we found it particularly valuable that both confirmed that the scores for individual YLLs reflected their own professional opinion of the students' level of vocabulary knowledge. It is also worth mentioning that the actual administration of YLVAT was easily done, and it did not require much time from the syllabus. Thus, a test such as YLVAT may be a useful tool for teachers in that it indeed seems to yield valid results at the individual level and therefore could be valuable for formative purposes. The scores on the separate parts may constitute starting points for individualized vocabulary work in the classroom. In line with the aim of the original VLT, namely to be used as a diagnostic tool (Nation, 2001), despite its summative characteristics, YLVAT has obvious formative qualities.

7. Concluding remarks and pedagogical implications

The need to investigate vocabulary knowledge among YLLs is evident, but so is the scarcity of suitable and relevant tests. Here we have presented YLVAT as a possible tool for teachers to gather details about their students' general L2 English vocabulary knowledge. Even though YLVAT can be used for summative purposes, its formative qualities must be emphasized considering the age of the intended test-takers. Thus, YLVAT can play an important role in EFL classrooms. As has been shown in this paper, the test has the capacity to reflect individual learners' general L2 English vocabulary knowledge; therefore, it is likely that YLVAT can facilitate teachers' identification of strong as well as weak learners in this regard, and from there classroom activities and vocabulary tasks can be adjusted. Thus, the findings presented in this paper have clear pedagogical implications. Further, as verified by the comments made by the participating teachers, YLVAT is easily administered and does not steal much time from valuable, regular lessons. Moreover, although perceived by the test-takers as a rather difficult vocabulary test, YLVAT was also considered fairly fun to take. In light of the fact that it is aimed at YLLs, the latter finding is very important since emotional growth in relation to vulnerability needs to be considered when children are involved (cf. McKay, 2006).

For a fuller understanding of the strengths and weaknesses of the test, YLVAT needs to be further investigated. For instance, the focus is on general vocabulary, rather than school-specific, and so YLVAT may work better in environments where extramural exposure to English is easily available. Therefore, it needs to be tested in different national context, and it would be necessary to substitute some of the cognates to suit the L1 of the intended task-takers. YLVAT should also be tested among many more YLLs and teachers, in combination with evaluation questions such as the ones used in the present study.

One part of YLVAT requires test-takers to produce lexical items. When working with these words, it was observed that learners differ greatly in terms of spelling ability. Although spelling may not be the most prioritized skill among YLLs, demands on accuracy grow with age, and it

would therefore be of interest to investigate spelling proficiency as demonstrated in YLVAT to see what characteristics, if any, are typical for good spellers already among YLLs.

In sum, YLVAT has been shown to possess many useful characteristics for determining YLLs' level of general vocabulary knowledge, and thus also paves the way for individually tailor-made classroom activities. In the future, we hope to see studies investigating YLVAT further, both quantitatively and qualitatively in various national contexts, in order to ascertain its full potential.

References

- Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, 24(4), 425-438. doi:10.1093/applin/24.4.425
- Aron, A., Aron, E. N., & Coups, E. J. (2005). *Statistics for the behavioral and social sciences: A brief course* (3 ed.). London: Prentice Hall International.
- Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing*, 27(1), 101-118.
- Berns, M., de Bot, K., & Hasebrink, U. (2007). *In the presence of English: Media and European youth*. New York, NY: Springer.
- Cambridge English. (2016). Retrieved from <http://www.cambridgeenglish.org/exams/young-learners-english/>
- Carter, R. (1987). *Vocabulary - applied linguistic perspectives*. London: Allen & Unwin.
- Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test manual (PPTVTM-4) (4 ed.)*. Minneapolis, MA: Pearson.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.
- Enever, J. (2011). *ELLiE: Early language learning in Europe*. Manchester: The British Council.
- Erickson, G., & Gustafsson, J.-E. (2005). *Some European Students' and Teachers' Views on Language Testing and Assessment*. Retrieved from EALTA webpage: www.ealta.eu.org > Resources
- Forsman, L. (2004). *Language, culture and context: Exploring knowledge and attitudes among Finland-Swedish EFL-students with particular focus on extracurricular influence*. (Lic.), Åbo Akademi, Vasa.
- García Mayo, M. d. P., & García Lecumberri, M. L. (2003). *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters.
- Hasselgreen, A. (1996). *Kartlegging av kommunikativ kompetanse i engelsk. User compendium for teachers*. Oslo: Nasjonalt læremiddelsenter.
- Hasselgreen, A., & Drew, I. (2012). Introduction. In A. Hasselgreen, I. Drew, & B. Sørheim (Eds.), *The young language learner: Research-based insights into teaching and learning*, pp. 5-8. Bergen: Fagbokforlaget.
- Hasselgreen, A., Drew, I., & Sørheim, B. (2012). *The young language learner: Research-based insights into teaching and learning*. Bergen: Fagbokforlaget.
- Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, 21(2), 303-317.
- Henry, A. (2013). Digital games and ELT: Bridging the authenticity gap. In E. Ushioda (Ed.), *International perspectives on motivation: Language learning and professional challenges*, pp. 133-155. Basingstoke: Palgrave Macmillan.
- Herriman, J. (1997). *Vocabulary and reading comprehension*. Göteborg: Göteborgs universitet.
- Huhta, A. (1996). Validating an EFL C-test for students of English philology. In R. Grotjahn (Ed.), *Der C-Test: theoretische Grundlagen und praktische Anwendungen [The C-test:*

- theoretical foundations and practical applications*] (Vol. 3, pp. 197–234). Bochum: Brockmeyer.
- Klein-Braley, C. (1996). Hunting unicorns: C-Tests, cloze tests and the RIP (Removal of Information Procedure). In R. Grotjahn (Ed.), *Der C-Test: theoretische Grundlagen und praktische Anwendungen* [*The C-test: theoretical foundations and practical applications*] (Vol. 3, pp. 127–137). Bochum: Brockmeyer.
- Laufer, B. (1995). Beyond 2000: A measure of productive lexicion in second language. In L. Eubank, L. Selinker, & M. Sharwood-Smith (Eds.), *The current state of interlanguage: Studies in honor of William E. Rutherford*, pp. 265-272. Amsterdam: Benjamins.
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics*, 19(2), 255-271.
- Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16(1), 33-51.
- Lextutor. (2016). Retrieved from www.lex tutor.ca
- McKay, P. (2006). *Assessing young language learners*. Cambridge: Cambridge University Press.
- McLean, S., Kramer, B., & Beglar, D. (2015). The validation of a listening vocabulary levels test. *Language Teaching Research*, 19(6), 741-760.
- Meara, P. (1992). *EFL vocabulary tests*. Swansea: Centre for applied language studies, University College Swansea.
- Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Performance & competence in second language acquisition*, pp. 35-53. Cambridge: Cambridge University Press.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (7th edition ed.). New York: Macmillan.
- Mochida, A., & Harrington, M. (2006). The Yes/No test as a measure of receptive vocabulary knowledge. *Language Testing*, 23(1), 73-98.
- Nation, P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nikolov, M. (2009). Introduction. In M. Nikolov (Ed.), *Early learning of modern foreign languages: Processes and outcomes*, pp. xiii-xvi. Bristol: Multilingual Matters.
- Nikolov, M. (2009). *Early learning of modern foreign languages: Processes and outcomes*. Bristol: Multilingual Matters.
- Olsson, E. (2011). *"Everything I read on the Internet is in English" - On the impact of extramural English on Swedish 16-year-old pupils' writing proficiency*. (Licentiate), University of Gothenburg, Gothenburg.
- Paribakht, T. S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition*, pp. 174-200. Cambridge: Cambridge University Press.
- Read, J., & von Randow, J. (2013). A university post-entry English language assessment: Charting the changes. *International Journal of English Studies*, 13(2), 89-110.
- Reuterberg, S.-E. (1999). Textinnehåll och könsskillnader i delprovet engelsk läsförståelse (ELF). In E. Johansson (Ed.), *Fokus på högskoleprovet* (Vol. 1999:6, pp. 36-42). Stockholm: Högskoleverket.

- Roessingh, H., & Elgie, S. (2009). Early language and literacy development among young English language learners: preliminary findings from a longitudinal study. *TESL Canada Journal*, 26(2), 24-45.
- Rogers, J., Webb, S., & Nakata, T. (2015). Do the cognacy characteristics of loanwords make them more easily learned than noncognates? *Language Teaching Research*, 19, 9-27. doi:10.1177/1362168814541752
- Scott Langeland, A. (2012). Investigating vocabulary development in English from grade 5 to grade 7 in a Norwegian primary school. In A. Hasselgreen, I. Drew, & S. B. (Eds.), *The young language learnign: Research-based insights into teaching and learning*, pp. 131-143. Bergen: Fagbokforlaget.
- Sundqvist, P. (2011). A possible path to progress: Out-of-school English language learners in Sweden. In P. Benson & H. Reinders (Eds.), *Beyond the language classroom*, pp. 106-118. Basingstoke: Palgrave Macmillan.
- Sundqvist, P., & Sylvén, L. K. (2012). World of VocCraft: Computer games and Swedish learners' L2 vocabulary. In H. Reinders (Ed.), *Digital games in language learning and teaching* (pp. 189-208). Basingstoke: Palgrave Macmillan.
- Sundqvist, P., & Sylvén, L. K. (2014). Language-related computer use: Focus on young L2 English learners in Sweden. *ReCALL*, 26(1), 3-20. doi:10.1017/S0958344013000232
- Sylvén, L. K., & Sundqvist, P. (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *ReCALL*, 24(3), 302-321. doi:10.1017/S095834401200016X
- The Government of the Netherlands. (2016). Retrieved from www.government.nl/issues/education/primary-education
- The Swedish National Agency for Education. (2004). *National evaluation of the compulsory school in 2003: A summary main report (A summary of report 250)*. Retrieved from Stockholm: www.skolverket.se
- The Swedish National Agency for Education. (2016). Retrieved from www.skolverket.se/bedomning
- Uzun, L. (2009). An evaluative checklist for computer games used for foreign language vocabulary learning and practice: VocaWord sample. *Novitas-ROYAL*, 3(1), 45-59.
- van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? . *Applied Linguistics*, 34(4), 457-479. doi:10.1093/applin/ams07
- Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, 30(1), 79-95.
- Webb, S., & Rodgers, M. P. H. (2009). The lexical coverage of movies. *Applied Linguistics*, 30(3), 407-427.

Appendix 1.
Young Learner Vocabulary Assessment Test (YLVAT)

VOCABULARY TEST, GRADE 6
ORDPROV, ÅRSKURS 6

First and last name/För- och efternamn: _____

Class/Klass: _____

School/Skola: _____

The test consists of three parts. Read the instructions carefully. If you don't understand what to do, please ask your teacher to explain! Good luck – and thanks for doing this.

Provet består av tre delar. Läs instruktionerna noga. Om du inte förstår vad du ska göra, be din lärare att förklara! Lycka till – och tack på förhand för att du gör detta.

Pia Sundqvist, fil. dr., Karlstads universitet
 Liss Kerstin Sylvén, fil. dr., Göteborgs universitet

PART A (Del A) – WORD RECOGNITION (Att känna igen ord)

Instructions / Instruktioner

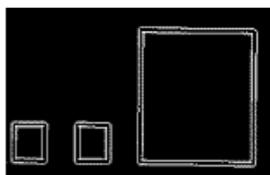
There are 13 questions in Part A. Check "T" if a sentence is **true**. Check "N" if a sentence is **not true**. Check "X" if you **do not understand** the sentence. Follow the example!

Del A består av 13 frågor. Kryssa för "T" om meningen är sann (True). Kryssa för "N" om meningen inte är sann (Not true). Kryssa för "X" om du inte förstår meningen. Följ exemplet!

Example) We can stop time.

- ☐ T (This is **True**)
☒ N (This is **Not true**)
☐ X (I **do not understand** the question)

1)



Two of these are little.

- ☐ T
☐ N
☐ X

2)

When someone says 'What are you called?', you should say your name.

- ☐ T
☐ N
☐ X

Figure 2. Vocabulary test, page 1.

3)	There are many ways to get money.	<input type="checkbox"/>	T
		<input type="checkbox"/>	N
		<input type="checkbox"/>	X
4)	All the world is under water.	<input type="checkbox"/>	T
		<input type="checkbox"/>	N
		<input type="checkbox"/>	X
5)	When you keep asking, you ask once.	<input type="checkbox"/>	T
		<input type="checkbox"/>	N
		<input type="checkbox"/>	X
6)	Sometimes people die when they fall off a building.	<input type="checkbox"/>	T
		<input type="checkbox"/>	N
		<input type="checkbox"/>	X
7)	Day follows night and night follows day.	<input type="checkbox"/>	T
		<input type="checkbox"/>	N
		<input type="checkbox"/>	X
8)	<i>Remain here</i> means 'stay'.	<input type="checkbox"/>	T
		<input type="checkbox"/>	N
		<input type="checkbox"/>	X
9)	When there is a change of scene, we see a different place.	<input type="checkbox"/>	T
		<input type="checkbox"/>	N
		<input type="checkbox"/>	X
10)	Dirty hands cannot make marks on glass.	<input type="checkbox"/>	T
		<input type="checkbox"/>	N
		<input type="checkbox"/>	X
11)	Each society has the same rules.	<input type="checkbox"/>	T
		<input type="checkbox"/>	N
		<input type="checkbox"/>	X
12)	Three examples of food are: shops, homes and markets.	<input type="checkbox"/>	T
		<input type="checkbox"/>	N
		<input type="checkbox"/>	X
13)	It is a short way from one side to the other side of a wide river.	<input type="checkbox"/>	T
		<input type="checkbox"/>	N
		<input type="checkbox"/>	X
PART A TOTAL SCORE:		<input type="text"/>	

Figure 3. Vocabulary test, page 2.

PART B (Del B) – MAKE THE RIGHT COMBINATIONS (Gör rätt kombinationer)**Instructions / Instruktioner**

There are 4 questions in Part B. Make the right combinations by writing the number next to the synonym or meaning of the word. Follow the example!

Del B består av 4 frågor. Gör rätt kombinationer genom att skriva siffran bredvid ordets synonym eller betydelse. Följ exemplet!

Example)	1.	original		
	2.	private	<u>6</u>	complete
	3.	royal	<u>1</u>	first
	4.	slow	<u>2</u>	not public
	5.	sorry		
	6.	total		
Question 1)	1.	apply		
	2.	elect	_____	choose by voting
	3.	jump	_____	become like water
	4.	manufacture	_____	make
	5.	melt		
	6.	threaten		
Question 2)	1.	blame		
	2.	hide	_____	keep away from sight
	3.	hit	_____	have a bad effect on something
	4.	invite	_____	ask
	5.	pour		
	6.	spoil		
Question 3)	1.	accident		
	2.	choice	_____	having a high opinion of yourself
	3.	debt	_____	something you must pay
	4.	fortune	_____	loud, deep sound
	5.	pride		
	6.	roar		
Question 4)	1.	basket		
	2.	crop	_____	money paid for doing a job
	3.	flesh	_____	heat
	4.	salary	_____	meat
	5.	temperature		
	6.	thread		

PART B TOTAL SCORE:

Figure 4. Vocabulary test, page 3.

PART C (Del C) – PRODUCTIVE VOCABULARY (Produktivt ordförråd)

Instructions / Instruktioner

Complete the underlined words as has been done in the example!

Skriv färdigt de understrukna orden på det sätt som exemplet visar!

Example) He was riding a bicycle.

- 1) Plants receive water from the soil through their ro.
- 2) The nu was helping the doctor in the operation room.
- 3) He is walking on the ti of his toes.
- 4) The mechanic had to replace the mo of the car.
- 5) There is a co of the original report in the file.
- 6) They had to cl a steep mountain to reach the cabin.
- 7) The railway con London with its suburbs.
- 8) The house was su by a garden.
- 9) This work is not up to your us standard.
- 10) She wan aimlessly in the streets.
- 11) They sat down to eat even though they were not hu.
- 12) The doctor ex the patient thoroughly.

PART C TOTAL SCORE:

TOTAL SCORE ON THE WHOLE TEST (Max: 37)

Figure 5. Vocabulary test, page 4.